Applets para SPSS – Manual de utilização

Autor: ENSP - Gabinete de Informática

Data: 18/06/2015

Objetivo:

Pré-requisitos: n/d

Controlo de versões:

Versão	Data	Autor
1.0	18/06/2015	João Tiago

Fonte: Manual de utilização da PSE Applets para SPSS (PSE – Produtos e Serviços de Estatística, Lda)

1 Introdução

O IBM SPSS Statistics, o software de análise estatística com maior quota de mercado não poderíamos deixar de colocar à disposição dos seus utilizadores o conhecimento acumulado da PSE (empresa distribuidora do IBM SPSS Statistics) de modo a facilitar processos de preparação e de análise de dados que, sendo possíveis de forma nativa no software, requerem tarefas mais ou menos complexas. Esta é a razão de existência da aplicação que designamos por 'PSE applets for SPSS'.

Nesta primeira versão incluímos 4 processos de preparação de dados:

- Junção de casos
- Junção de variáveis
- Extração de caracteres
- Criação de índices

Todos os processos são possíveis de realizar no IBM SPSS Statistics através de sintaxe, de múltiplas opções de menus ou pela repetição de um mesmo procedimento. Nos menus que criámos facilitamos a realização dos processos indicados – através de um interface único (uma caixa de diálogo) o utilizador faz as especificações do que pretende e obtém um resultado.
A cada menu está associado um botão de 'Help' onde pode aceder à informação sobre cada processo. Para além disso, criámos este manual em que de uma forma mais detalhada o ajudamos na utilização da aplicação 'PSE applets for SPSS'.
Esta aplicação estará disponível comercialmente seja para utilização dos atuais utilizadores SPSS, de acordo com a politica comercial definida pela PSE, seja para novos utilizadores.
Para novos utilizadores a PSE passa a disponibilizar a solução SPSS for Statistical Analysis que, para além de incluir o software IBM SPSS Statistics em qualquer configuração, incluirá também a aplicação 'PSE applets for SPSS'.
A aplicação 'PSE applets for SPSS' só pode ser utilizada com as versões de software IBM SPSS Statistics 21 ou superior.

2 Junção de Casos

A utilização deste procedimento tem as seguintes vantagens face àqueles que estão disponíveis nativamente com o IBM SPSS Statistics:

- Permite juntar informação arquivada até 50 ficheiros diferentes SPSS (.sav) através de um opção de menu e de uma só vez.
- Permite identificar o ficheiro de origem de cada caso presente no novo ficheiro de dados.
- O utilizador pode especificar a lista de variáveis a incluir no novo ficheiro de dados.
- Caso as variáveis especificadas não existam em todos os ficheiros, o utilizador é alertado através de mensagem no ficheiro de resultados do SPSS (output – .spv).

2.1 Funcionalidades

Este menu permite juntar informaçao proveniente de um mínimo de 2 até um máximo de 50 ficheiros SPSS (.sav) utilizando uma caixa de diálogo de fácil utilização.

Para executar o procedimento todos os ficheiros têm de estar no formato SPSS e, os ficheiros a serem juntos, disponíveis numa única pasta de dados. As designações das variáveis deverão ser as mesmas, por forma a que se consiga juntar diversos casos para uma mesma variável.

O utilizador pode selecionar, a partir do ficheiro principal, as variáveis para que pretende juntar informação. As restantes variáveis, não selecionadas, não serão consideradas no novo ficheiro de dados. Esta funcionalidade é extremamente útil quando o ficheiro principal, ou qualquer um dos outros ficheiros a juntar, contém um número elevado de variáveis e muitas delas não são de interesse para uma qualquer análise em que necessita de juntar informação distribuída.

Por exemplo, consideremos que é analista de vendas de uma empresa e que dispõe de um ficheiro de vendas para cada distrito. Esses ficheiros contêm informação diversa e normalizada em que, para além do volume de vendas, contém muita outra informação. Se o seu objetivo é exclusivamente o de prever a evolução do volume de vendas, é-lhe muito útil poder juntar a informação dos 18 ficheiros num único e somente com uma variável de data, uma outra de localização e outra de volume de vendas.

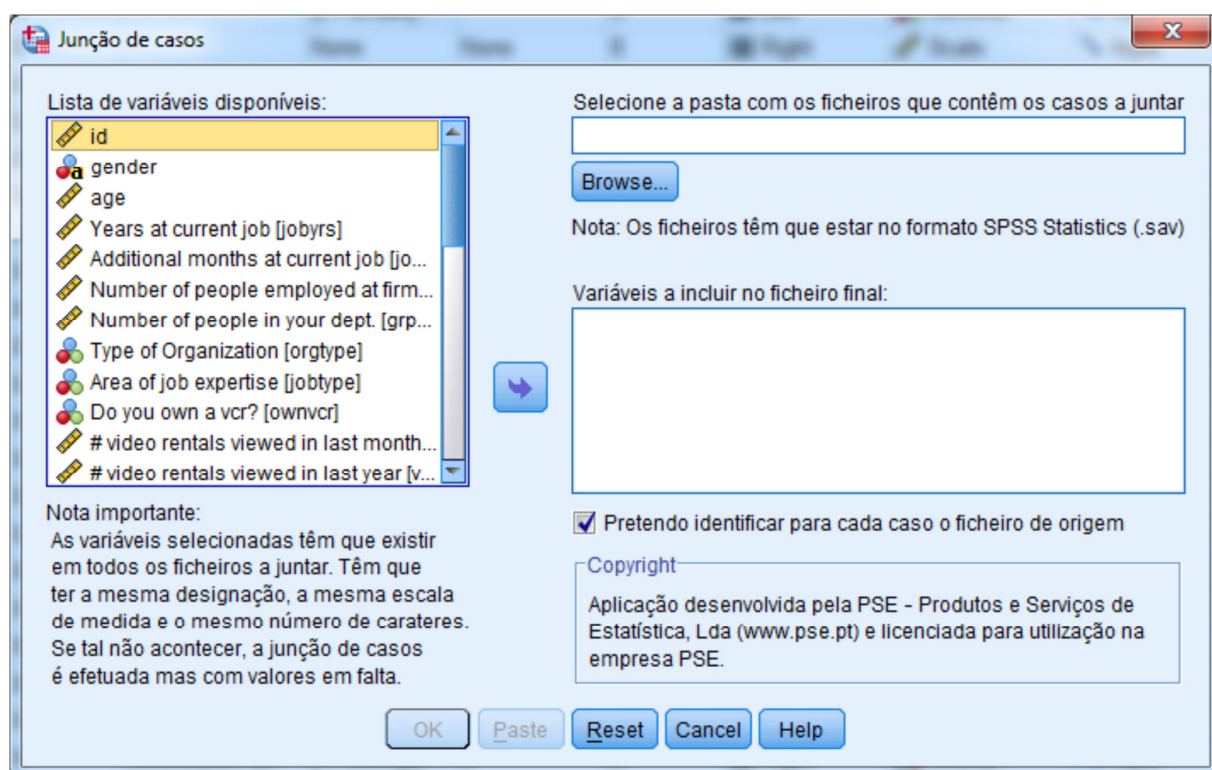
O resultado de junção é um novo dataset (ficheiro de dados) que contém todos os casos (anteriormente dispersos por todos os ficheiros) e as variáveis especificadas no procedimento. Os casos do ficheiro principal encontram-se no início do novo dataset, seguindo-se os restantes casos

dos ficheiros identificados na pasta de dados, definida pelo analista. No final do novo dataset é criada automaticamente uma variável que 7

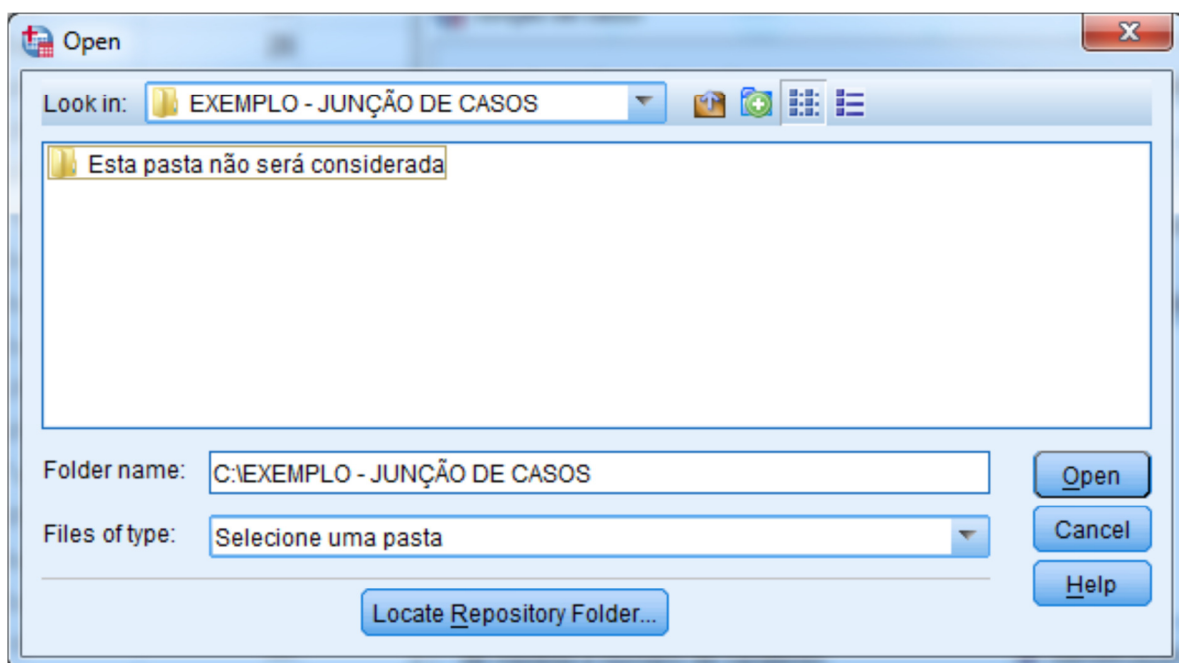
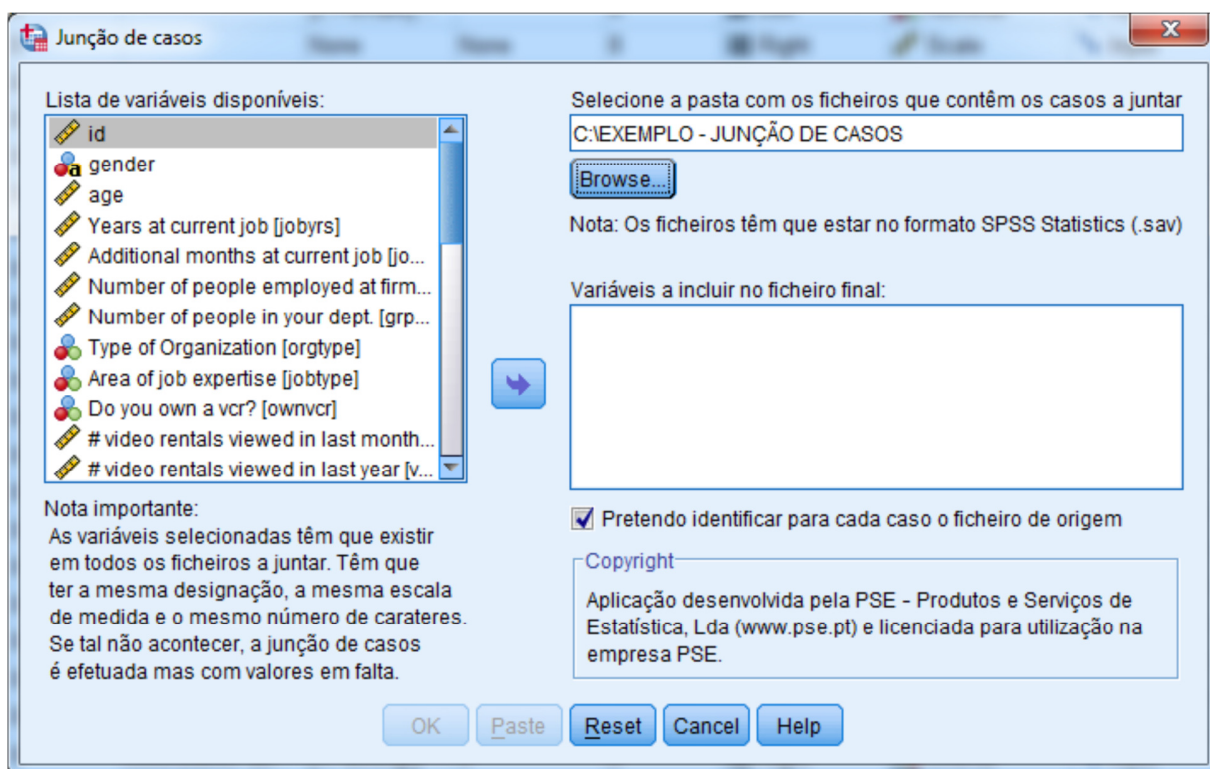
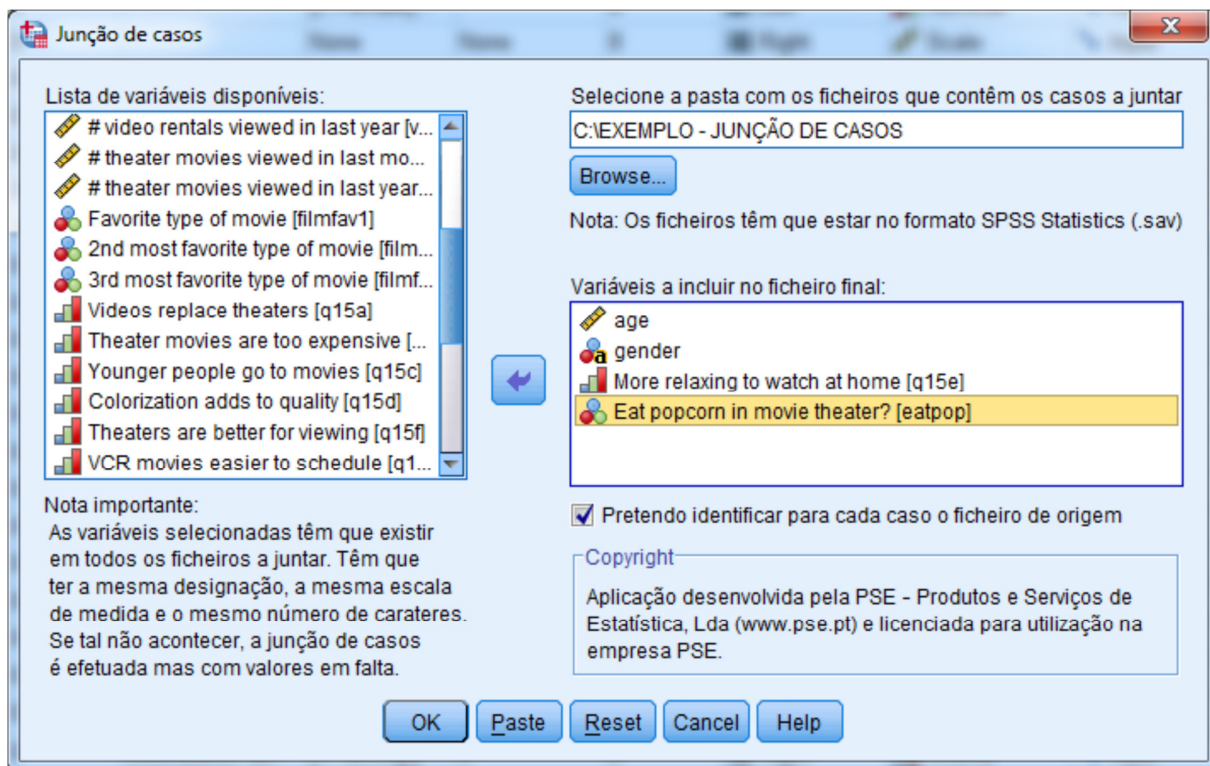
identifica o ficheiro de origem de cada caso. É uma variável numérica em que o descritivo de cada número é o nome do respetivo ficheiro de dados origem.

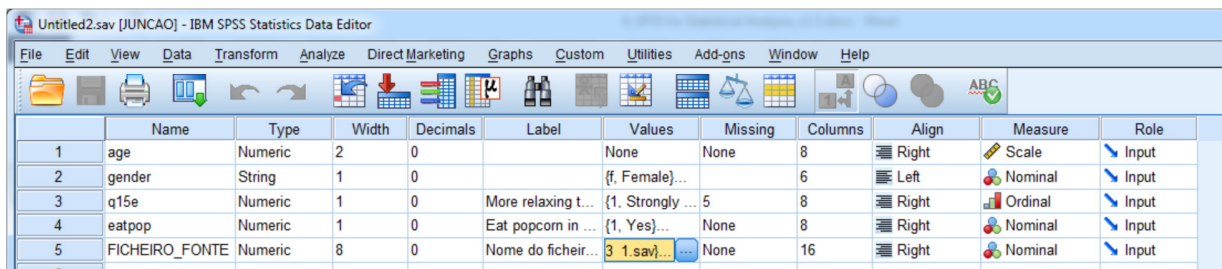
No ficheiro de resultados do SPSS (Output) são listados os ficheiros e as variáveis utilizadas no procedimento.

2.2 Utilização

Depois de abrir o ficheiro de dados principal, clique no menu **Custom** e selecione a opção **Junção de casos**. Aparece no ecrã a seguinte janela:

Na opção **Lista de variáveis disponíveis**, aparecerão as variáveis existentes no ficheiro principal. Por ficheiro principal, referimo-nos ao ficheiro a partir do qual serão identificadas as variáveis possíveis de utilizar na execução do procedimento.

Para executar a operação de junção tem que obrigatoriamente especificar a pasta onde se encontram os ficheiros a juntar com o ficheiro principal aberto. Clique no botão **Browse** e selecione a pasta que contém os ficheiros a juntar. Quando estiver dentro da pasta que pretende selecionar, clique no botão **Open**. O caminho completo é especificado na caixa **Folder name**.

Na imagem abaixo foi especificada a pasta "EXEMPLO - JUNÇÃO DE CASOS". Esta pasta inclui os ficheiros que pretendemos juntar. Apenas é considerado o conteúdo desta pasta, não o conteúdo das subpastas.

Após selecionar a opção **Open**, o caminho para a pasta de interesse encontrar-se-á já definido, conforme próxima imagem.

Para especificar as variáveis de interesse no ficheiro final, selecione as variáveis da **Lista de variáveis disponíveis** e arraste-as para a lista na parte direita da janela e denominada por **Variáveis a incluir no ficheiro final**.

Nota:

As variáveis utilizadas no procedimento de junção de casos podem assumir qualquer escala de medida (Nominal, Ordinal, Scale, Data e String (alfanuméricas). Variáveis de resposta múltipla não poderão ser utilizadas neste procedimento.

Em **Pretendo identificar para cada caso o ficheiro de origem** (seleção opcional), o analista tem a possibilidade de identificar a que ficheiro de dados correspondem os casos adicionados. A designação da nova variável corresponderá a **FICHEIRO_FONTE** e assumirá tantos valores quanto o número de ficheiros de dados utilizados na operação de junção. Cada valor terá como descritivo a designação de um ficheiro de dados.

Finalmente, são apresentados cinco botões para efeitos de execução:

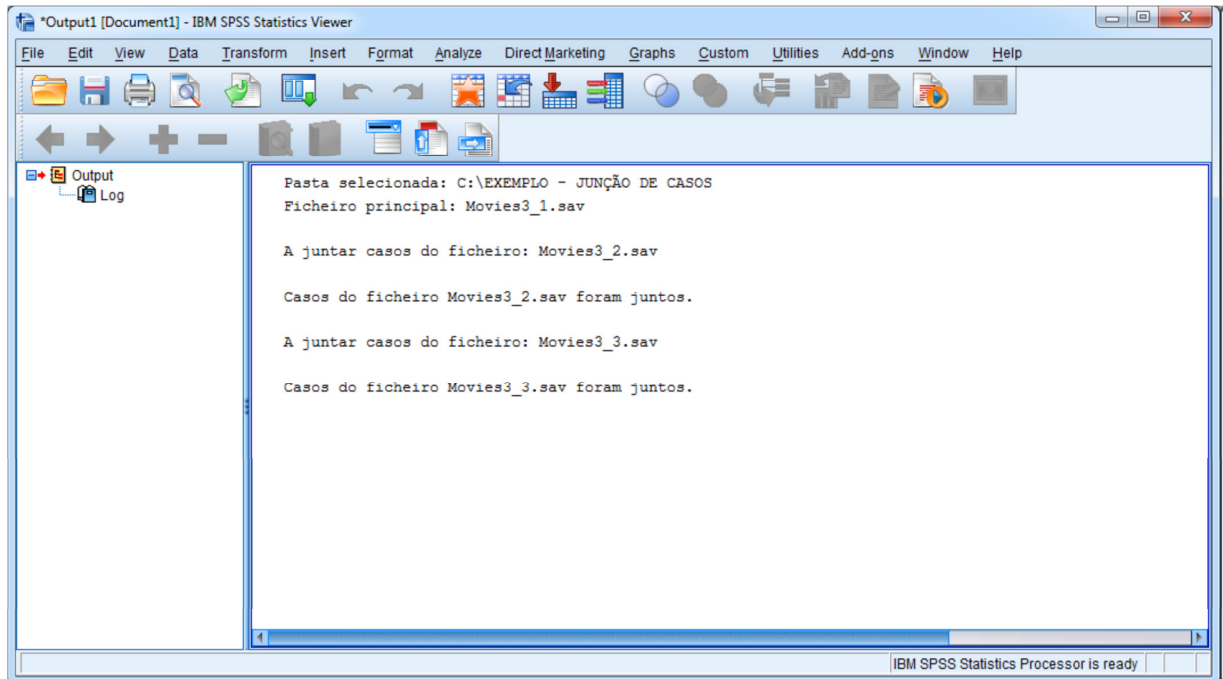
- ✓ **OK** – Corresponde à execução do procedimento. Esta opção tem como retorno a criação de dois novos ficheiros. Um ficheiro de dados cuja designação

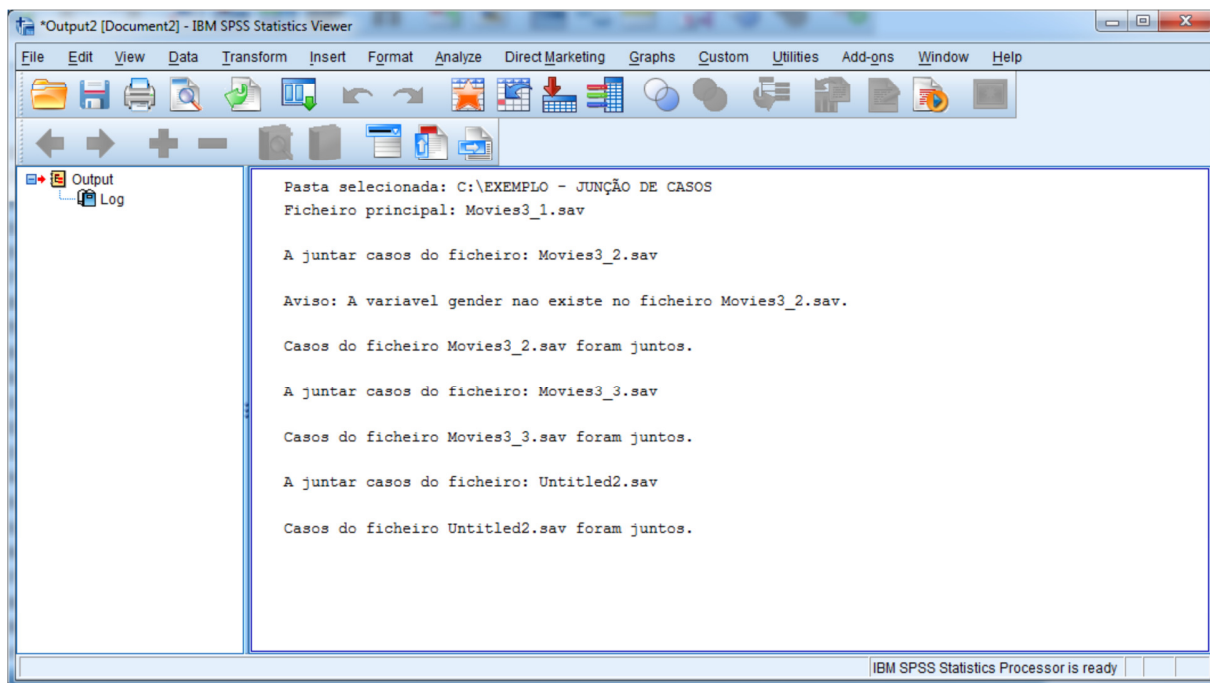
"**JUNCAO**" correspondente ao somatório dos casos de todos os ficheiros analisados; e um ficheiro de resultados – Output SPSS – contendo a identificação dos ficheiros de dados e das variáveis utilizadas na análise.

De seguida iremos analisar o novo dataset **JUNCAO**. Neste novo ficheiro de dados são incluídas apenas as variáveis que selecionámos para inclusão no ficheiro final. Uma vez que foi solicitada a criação de uma nova variável com identificação de ficheiro de origem, o ficheiro final contém uma nova variável FICHEIRO_FONTE. Esta variável identifica o ficheiro de origem de cada caso no dataset final.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	age	Numeric	2	0		None	None	8	Right	Scale	Input
2	gender	String	1	0		{f, Female}...		6	Left	Nominal	Input
3	q15e	Numeric	1	0	More relaxing t...	{1, Strongly ...	5	8	Right	Ordinal	Input
4	eatpop	Numeric	1	0	Eat popcorn in ...	{1, Yes}...	None	8	Right	Nominal	Input
5	FICHEIRO_FONTE	Numeric	8	0	Nome do ficheir...	3 1.sav}...	None	16	Right	Nominal	Input

O dataset JUNCAO contém todos os casos que existem nos ficheiros juntos (no nosso exemplo, existem na pasta 3 ficheiros).

Adicionalmente, será produzido o seguinte resultado na janela Output:

Temos informação sobre a pasta selecionada e a designação do ficheiro principal. Iterativamente é reportado o estado e o resultado de junção de casos de cada ficheiro.

Também aqui podemos reparar em avisos. Na próxima imagem, a variável gender não existe no ficheiro Movies3_2.sav. Neste caso, a junção é feita mas os casos da variável gender que são provenientes deste ficheiro estarão em falta (não preenchidos).

- ✓ **Paste** – Tal como acontece com os restantes procedimentos de menu do IBM SPSS Statistics, esta opção corresponde à apresentação dos comandos de programação associados ao procedimento de junção de casos. Esta janela de programação poderá ser guardada e utilizada conjuntamente com outros comandos de programação disponíveis no IBM SPSS Statistics.

Poderá partilhá-la com outros analistas, todavia esta partilha só terá sucesso na sua execução desde que o menu se encontre implementado no computador dos restantes utilizadores.

- ✓ **Reset** – Opção que permite limpar definições tidas no menu de junção de casos.
- ✓ **Cancel** – Opção que permite sair do procedimento de junção de casos.
- ✓ **Help** – Opção que permite aceder ao sistema de ajuda associado ao procedimento de junção de casos.

Junção de Casos

Terminologia

Casos
Registos em linha de um ficheiro de dados SPSS Statistics. Também conhecidos por observações.
Variáveis
Identificação das colunas na matriz de dados SPSS e que contém informação sobre cada um dos casos.
Dataset
Ficheiro de dados SPSS Statistics.
Lista de variáveis disponíveis
Identificação de variáveis disponíveis para seleção e consequente junção de casos.
- Só serão listadas variáveis provenientes do ficheiro principal.
- Em todos os ficheiros de dados as variáveis, identificadas para efeitos de junção, devem ter a mesma designação (Name), o mesmo tipo (Type) e o formato (Width e Decimals).
- Variáveis de resposta múltipla não podem ser especificadas.

Selecione a pasta com os ficheiros que contêm os casos a juntar
Identificação da pasta de diretoria Windows que contém os ficheiros a serem utilizados no procedimento.
- Os ficheiros têm de estar no formato SPSS Statistics (.sav).
- Subpastas, se existirem, não serão incluídas.
- Todos os ficheiros contidos na pasta serão utilizados na junção de casos.

Variáveis a incluir no ficheiro final
Identificação das variáveis a utilizar pelo procedimento.
- Só as variáveis especificadas constarão no ficheiro final.

Pretendo identificar para cada caso o ficheiro de origem (Pré-selecionado)
Criação de uma variável (designada FICHEIRO_FONTE) no dataset JUNCAO que permite identificar a que ficheiro de dados pertence cada caso. Caso não pretenda utilizar esta funcionalidade deve retirar a seleção existente.

Utilidade

Este procedimento permite juntar, num único ficheiro de dados, casos de uma ou mais variáveis que estejam distribuídos por múltiplos ficheiros de dados SPSS Statistics. O número de ficheiros de dados IBM SPSS Statistics a utilizar não deve ser superior a 50.

Funcionamento

Antes de iniciarmos o procedimento devemos colocar os ficheiros de dados SPSS, onde estão os casos, numa pasta de diretoria Windows.
Esta pasta será posteriormente identificada durante a execução do procedimento e colocada na caixa respetiva.
Na caixa 'Lista de variáveis disponíveis' devemos selecionar as variáveis cujos casos queremos juntar.
Utilizando o botão com a seta devemos colocá-las na caixa 'Variáveis a incluir no ficheiro final'.

Considerações

- O dataset principal tem que ser um ficheiro SPSS arquivado numa pasta de dados. Este ficheiro pode estar arquivado na mesma pasta que os outros ficheiros a juntar.
- As variáveis têm que conter as mesmas designações (Name), o mesmo tipo (Type) e o mesmo formato (Width e Decimals) em todos os ficheiros a juntar. Caso contrário, a junção falha.
- As variáveis que apresentam o mesmo conteúdo mas as suas designações são diferentes (por exemplo, género e sexo) são tratadas como variáveis independentes. Estas variáveis apenas têm valores preenchidos no bloco de casos provenientes do ficheiro de origem. Para os restantes ficheiros os casos aparecerão preenchidos como *system missings*.
- Caso não existam ficheiros a juntar, o comando processa apenas o ficheiro principal (incluindo apenas as variáveis selecionadas).
- Sugerimos que, antes de proceder à junção de casos, mantenha apenas o ficheiro principal aberto. Os restantes deverão manter-se fechados na respetiva pasta de dados.

3 Junção de variáveis

Este procedimento permite juntar variáveis a um ficheiro de dados com origem de múltiplos ficheiros SPSS – de um mínimo de 2 a um máximo de 50 - através de um interface gráfico e de uma única vez. No IBM SPSS Statistics, através da opção de menu Data – Merge Files – Add Variables, só é possível juntar variáveis de 1 ficheiro de dados em cada operação.

3.1 Funcionalidades

Para executar o procedimento todos os ficheiros a incluir no procedimento devem estar no formato SPSS (.sav) e colocados numa única pasta de dados.

Existem duas opções no que respeita ao processo de junção de variáveis - Com ou sem especificação de variável(eis) chave.

> *Aconselhamos a que só proceda a esta operação utilizando pelo menos uma variável chave que, devendo ser unívoca, garante a correta inclusão da informação das variáveis a juntar nos casos existentes no designado ficheiro principal.*

Ao especificar uma ou mais variáveis chave (obrigatoriamente comuns a todos os ficheiros de dados a utilizar), estas deverão ter a mesma designação e o mesmo tipo em todos os ficheiros.

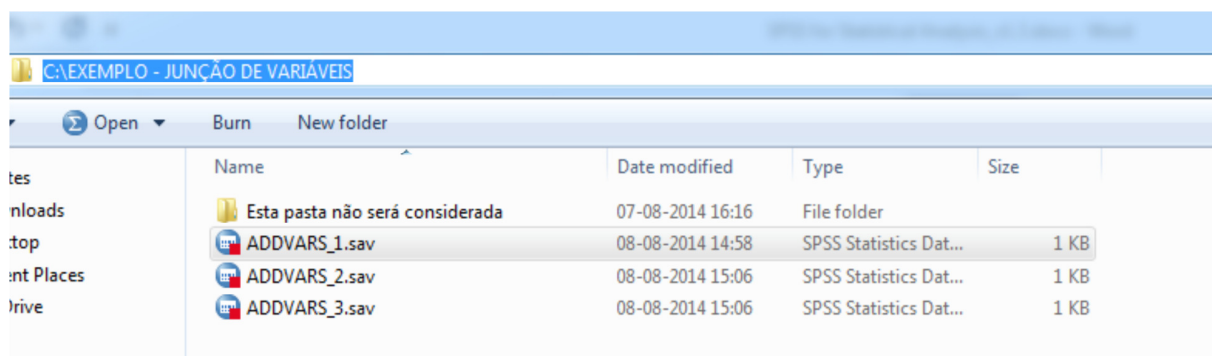
A opção de menu de junção de variáveis permite identificar a forma como a informação em todos os ficheiros será identificada univocamente (Variáveis chave), ou seja, como serão ligados os diferentes ficheiros com o ficheiro principal a cujos casos serão juntos a informação contida nas novas variáveis a juntar.

Ao executar este procedimento obterá dois resultados - um novo dataset que irá conter todos os casos e as variáveis do ficheiro principal juntamente com as variáveis identificadas para junção e provenientes dos restantes ficheiros de dados. Adicionalmente, um ficheiro de resultados SPSS (Output) que identificará os ficheiros e as variáveis juntas e possíveis inconsistências que possam ocorrer. (*Por exemplo, se estiver a utilizar uma variável chave que, sendo do tipo alfanumérico, apresente o campo "Width" diferente entre os diversos ficheiros, será reportada uma inconsistência*).

3.2 Utilização

Nota:

Para exemplificar uma aplicação prática desta opção de menu foram criados 3 ficheiros de dados - ADDVARS_1.sav (que corresponderá ao ficheiro principal), ADDVARS_2.sav e ADDVARS_3.sav.

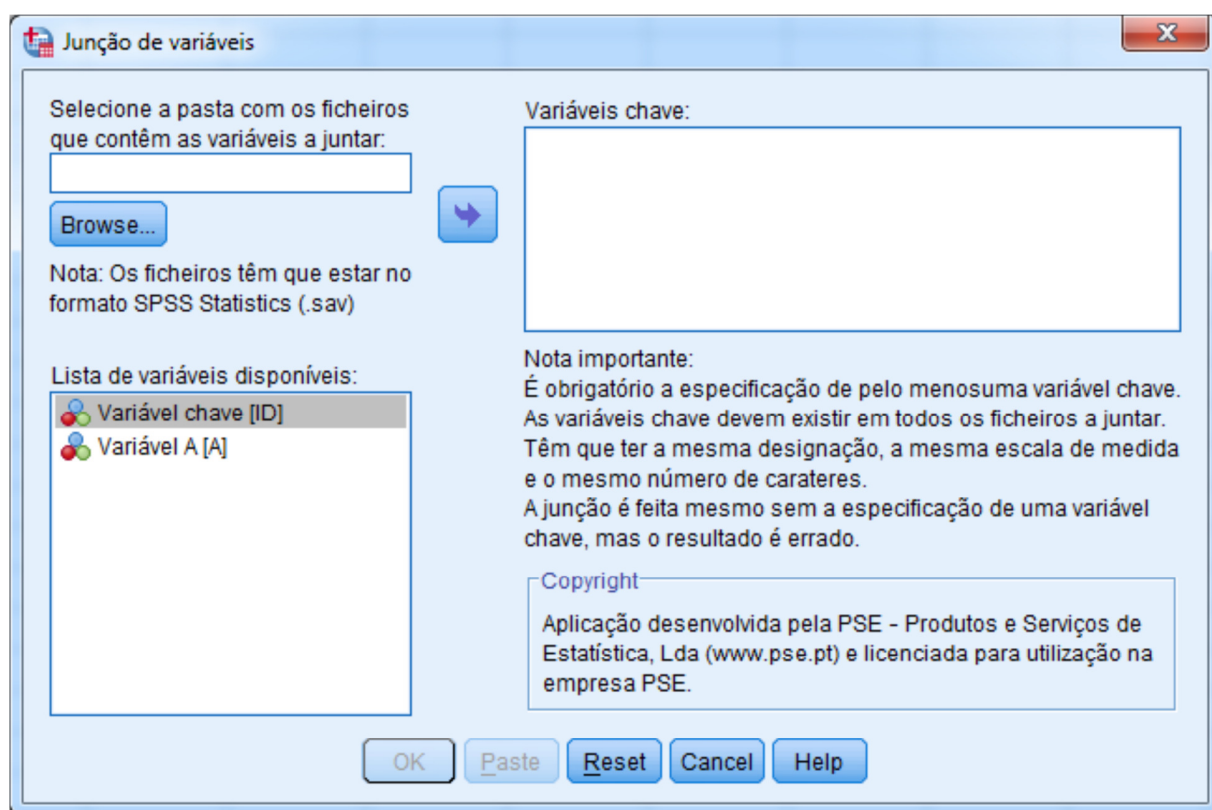
Os dois últimos ficheiros encontram-se numa pasta com o seguinte caminho:

A estrutura de cada ficheiro é a seguinte:

ADDVARS_1.sav	ADDVARS_2.sav	ADDVARS_3.sav
Variáveis		
1. ID	1. ID	1. ID

2. *A	2. B	2. C

onde a variável ID corresponderá à variável chave (isto é a variável que permitirá associar univocamente os casos de todos os ficheiros). O ficheiro final conterá as variáveis "ID", "A", "B" e "C", mantendo para cada caso a respetiva correspondência entre ficheiros.

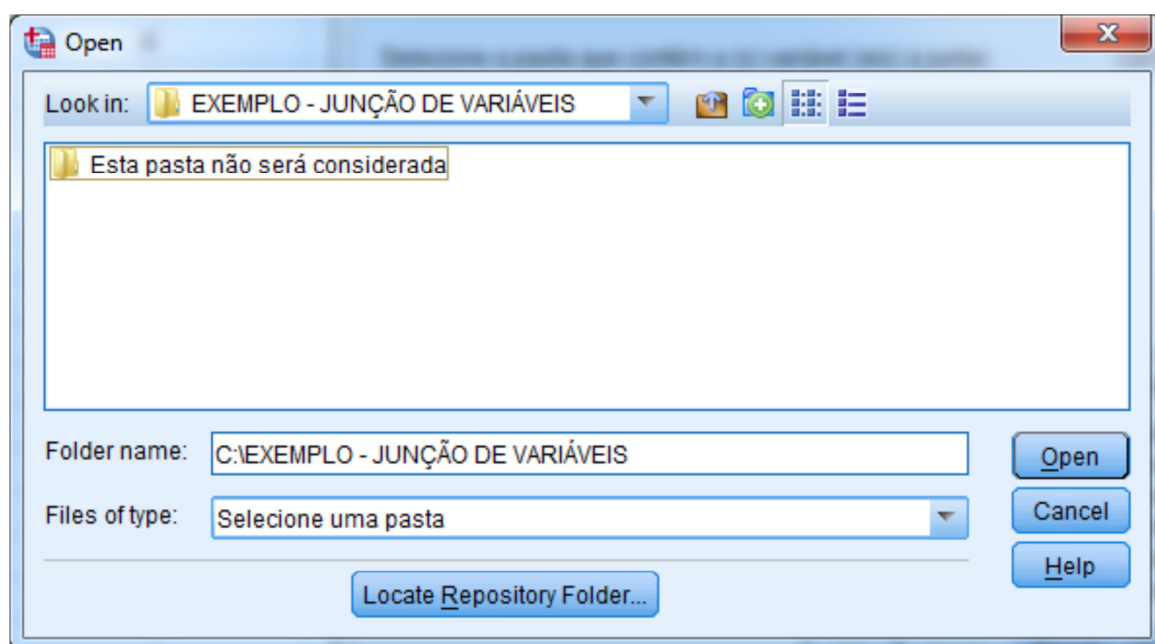
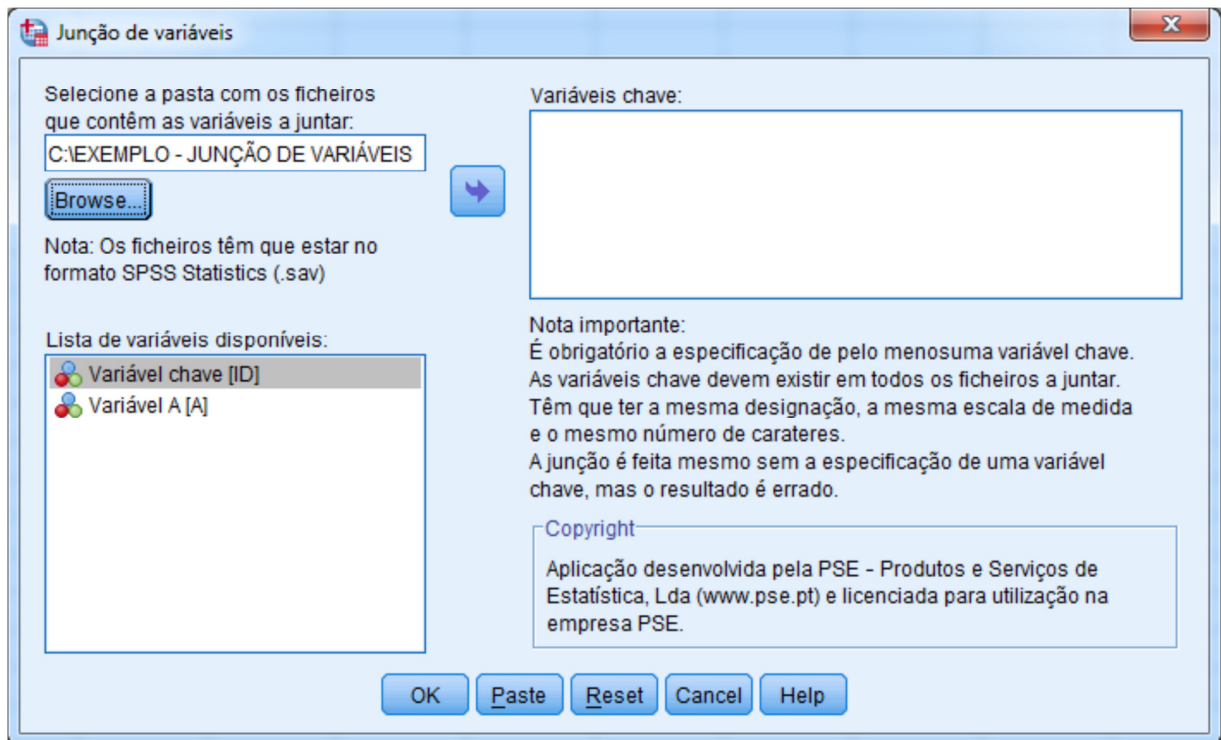
Depois de abrir o ficheiro de dados, clique no menu **Custom** e selecione a opção **Junção de variáveis**.

No ecrã aparecerá uma janela igual à da imagem seguinte:

Para executar a junção de variáveis tem que obrigatoriamente especificar a pasta onde se encontram os ficheiros a juntar ao ficheiro principal (que está aberto).

Clique no botão **Browse** e selecione a pasta que contem os ficheiros a juntar.

Quando estiver dentro da pasta que pretende selecionar, clique no botão **Open**

O caminho completo é especificado na caixa **Folder name**.

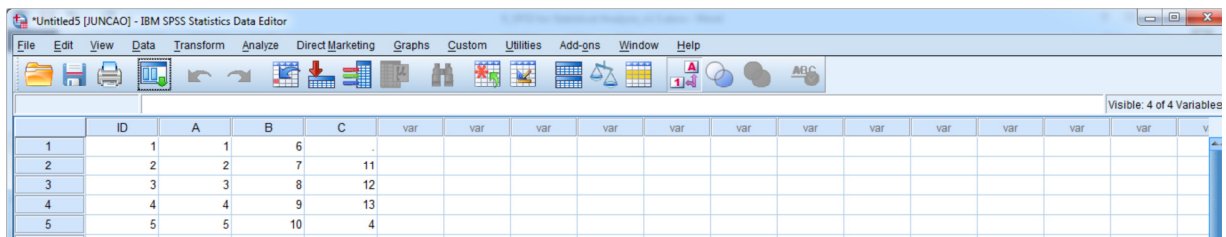
Na imagem abaixo foi especificada a pasta “EXEMPLO - JUNÇÃO DE VARIÁVEIS”. Esta pasta inclui os ficheiros que pretendemos juntar. Apenas é considerado o conteúdo desta pasta, não o conteúdo das subpastas.

Após selecionar a opção **Open**, o caminho para a pasta encontrar-se-á já definido, conforme próxima imagem.

A partir do momento em que o caminho de dados se encontra especificado, já será possível executar o procedimento de junção de variáveis com os restantes ficheiros de dados, sem especificar quaisquer variáveis chave.

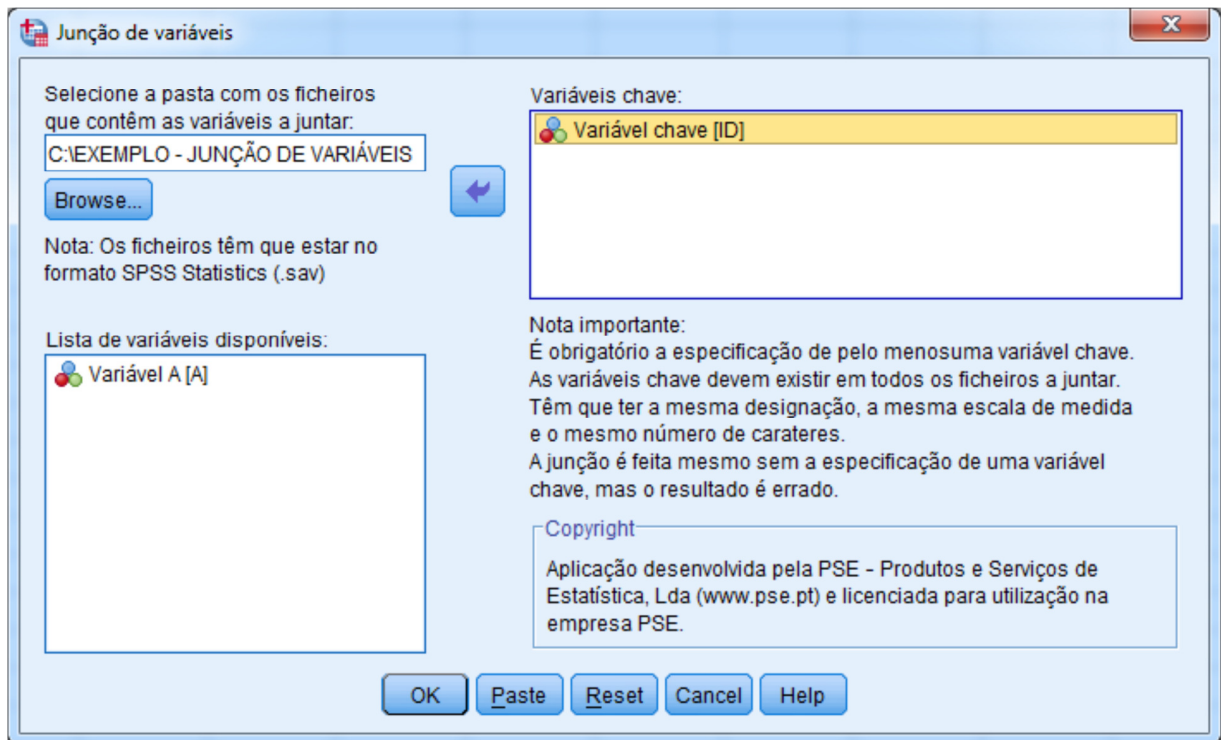
Se proceder como indicado, o resultado será o seguinte:

- É gerado um novo dataset gerado com a designação "**JUNCAO**", onde os casos serão juntos de forma sequencial em cada variável. Nada garante que a informação relativa ao caso com o ID 1 seja igual à informação para este ID em cada um dos 3 ficheiros. Tal só ocorrerá se os 3 ficheiros estiverem ordenados da mesma forma pela variável ID.

	ID	A	B	C	var	var	var	var	var	var	var
1	1	1	6	11							
2	2	2	7	12							
3	3	3	8	13							
4	4	4	9	4							
5	5	5	10	.							
6											

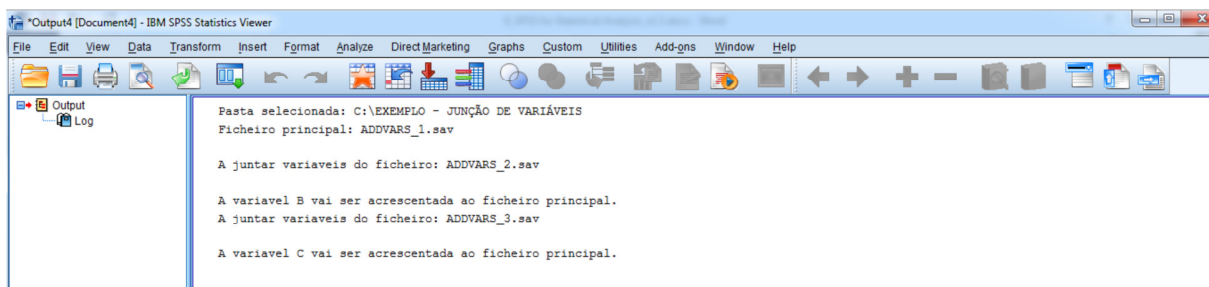
- No ficheiro de resultados SPSS (Output) são apresentadas as variáveis e os ficheiros incluídos no processamento (de referir que não são identificadas inconsistências, apesar de podermos por em causa a consistência da informação para cada caso).

Para evitar quaisquer problemas de inconsistência da informação para cada caso, na caixa **Variáveis chave** deve especificar a(s) variável(eis) que constituirão a chave únivoca de ligação entre os ficheiros de dados.

Da caixa "**Lista de variáveis disponíveis**" deve selecionar as variáveis a incluir como variáveis chave e com o botão colocá-las na caixa "**Variáveis chave**".

Retomamos o nosso exemplo, desta vez especificando a variável "ID" em Variáveis chave.

O resultado final corresponderá, tal como anteriormente, à criação de dois ficheiros:

- Um novo dataset com a designação “**JUNCAO**”, onde a informação constante para cada caso corresponde à informação que cada ID tinha nos ficheiros iniciais.

	ID	A	B	C
1	1	1	6	.
2	2	2	7	11
3	3	3	8	12
4	4	4	9	13
5	5	5	10	4

- Tal como antes, no ficheiro de resultados SPSS (Output) são apresentadas as variáveis e os ficheiros incluídos no processamento.

Finalmente, são apresentados cinco botões para efeitos de execução:

- ✓ **OK** – Para execução imediata do procedimento e obtenção dos resultados.
- ✓ **Paste** – Tal como acontece com os restantes procedimentos de menu do IBM SPSS Statistics, esta opção corresponde à apresentação dos comandos de programação associados ao procedimento de junção de variáveis. Esta janela de programação poderá ser guardada e utilizada conjuntamente com outros comandos de programação disponíveis no IBM SPSS Statistics.

- ✓ **Reset** – Opção que permite limpar definições tidas no menu de junção variáveis.
- ✓ **Cancel** – Opção que permite sair do procedimento de junção de variáveis.
- ✓ **Help** – Opção que permite aceder ao sistema de ajuda associado ao procedimento de junção de variáveis.

C:\Program Files\IBM\SPSS\Statistics\21\ext\lib\menu2\menu2.htm

Junção de Variáveis

Terminologia

Casos
: Registos em linha de um ficheiro de dados SPSS Statistics. Também conhecidos por observações.

Variáveis
: Identificação das colunas na matriz de dados SPSS e que contém informação sobre cada um dos casos.

Dataset
: Ficheiro de dados SPSS Statistics.

Lista de variáveis disponíveis
: Identificação de variáveis disponíveis para seleção como Variáveis Chave.
- Só serão listadas variáveis provenientes do ficheiro principal.
- Em todos os ficheiros de dados as variáveis, selecionadas como variáveis chave, devem ter a mesma designação (Name), o mesmo tipo (Type) e o mesmo formato (Width e Decimals).
- Variáveis de resposta múltipla não podem ser especificadas.

Selecione a pasta com os ficheiros que contêm as variáveis a juntar
: Identificação da pasta de diretoria Windows que contém os ficheiros a serem utilizados no procedimento.
- Os ficheiros têm de estar no formato SPSS Statistics (.sav).
- Subpastas, se existirem, não serão incluídas.
- Todos os ficheiros contidos na pasta serão utilizados na junção de variáveis.

Variáveis chave
: Identificação das variáveis que irão condicionar a operação de junção de novas variáveis de modo a garantir que cada registo das novas variáveis é incluído corretamente no novo ficheiro de dados.
- Pode selecionar uma ou mais variáveis como variáveis chave
- O importante é que cada registo do ficheiro de dados permita ser identificado univocamente. Ou seja, que não existam casos duplicados e que cada um seja identificado de forma diferente pela conjugação das variáveis chave.

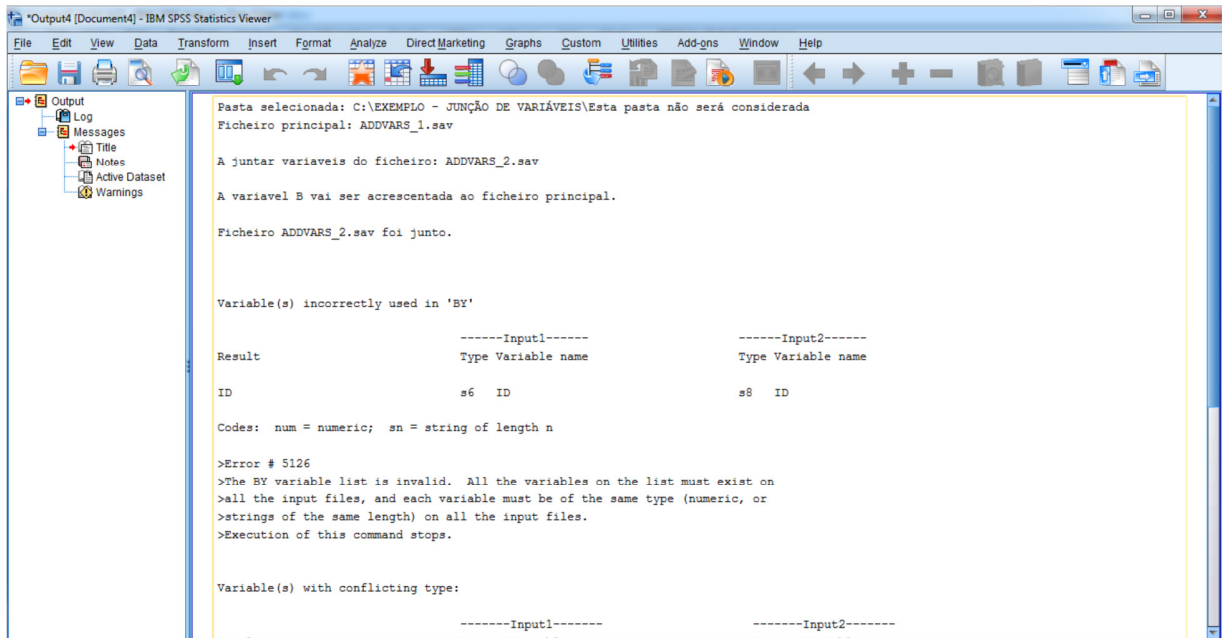
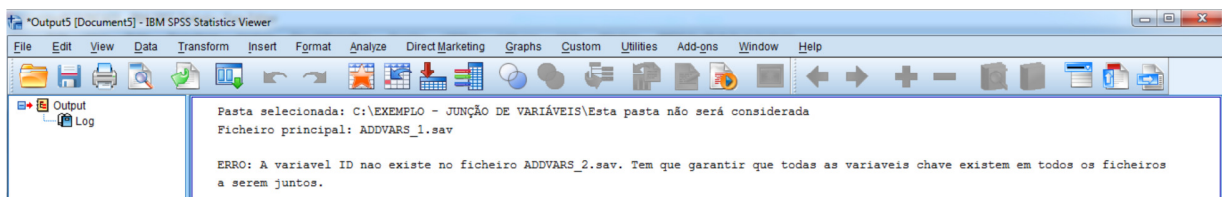
Utilidade

Este procedimento permite juntar, num único ficheiro de dados, variáveis relativas aos mesmos registos que estejam repartidas por múltiplos ficheiros de dados SPSS Statistics. O número de ficheiros de dados IBM SPSS Statistics a utilizar não deve ser superior a 50.

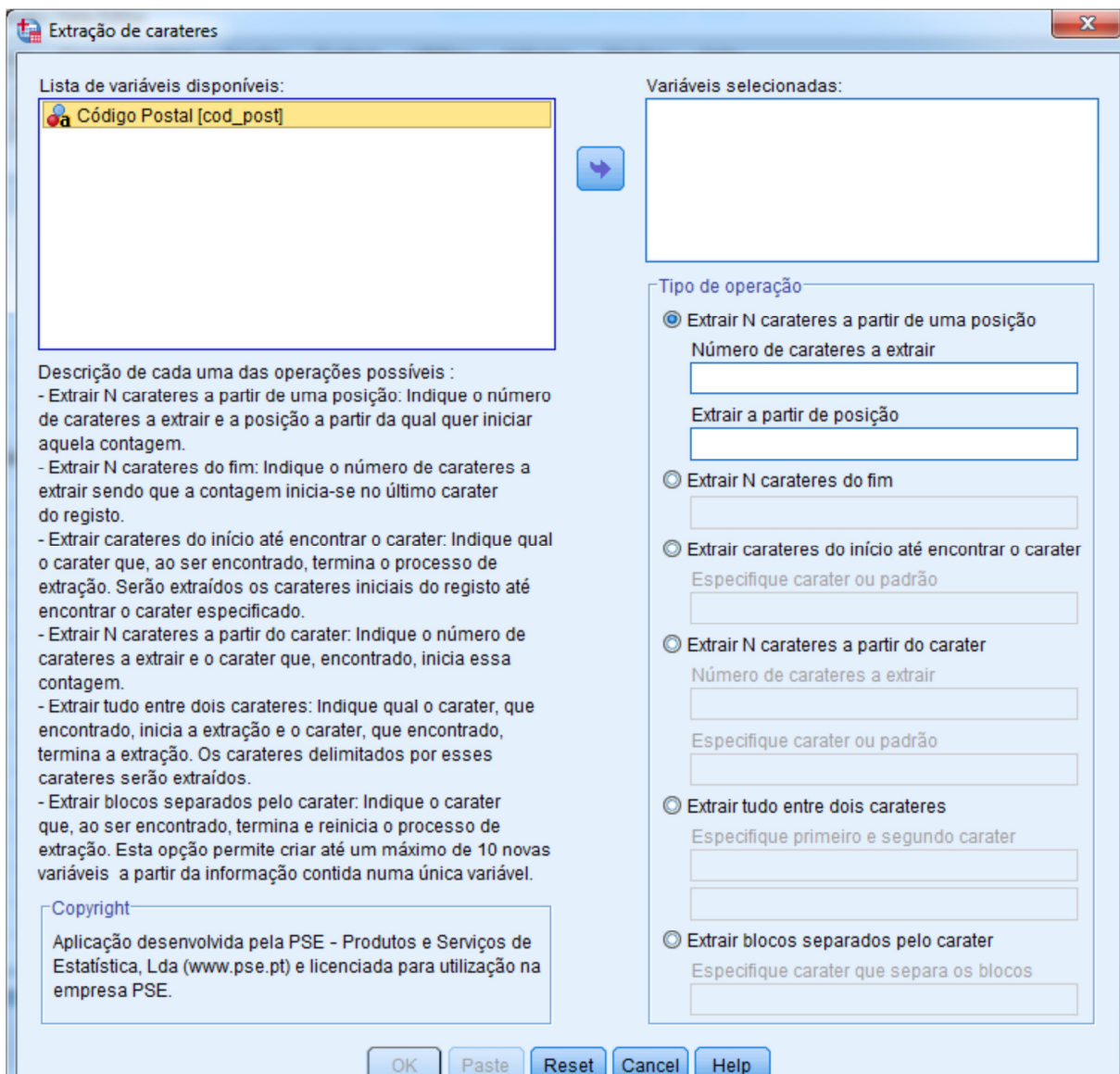
Funcionamento

Antes de iniciarmos o procedimento devemos colocar os ficheiros de dados SPSS Statistics, onde estão as variáveis, numa pasta de diretoria Windows.
Esta pasta será posteriormente identificada durante a execução do procedimento e colocada na caixa respetiva.
Na caixa 'Lista de variáveis disponíveis' devemos selecionar as variáveis cuja combinação permitem criar uma chave unívoca para cada caso.
Utilizando o botão com a seta devemos colocá-las na caixa 'Variáveis chave'.
Escolhendo o botão OK executaremos de imediato o procedimento, cujos resultados serão os seguintes:

Considerações

- O dataset principal tem de ser um ficheiro SPSS arquivado numa pasta de dados. Este ficheiro pode estar arquivado na mesma pasta que os outros ficheiros a juntar.
- Quando se especifica uma (ou mais) variáveis chave, estas têm que conter as mesmas designações (Name), o mesmo tipo (Type) e o mesmo formato (Width e Decimals) em todos os ficheiros a juntar. Caso contrário, no ficheiro de resultados SPSS (Output) será feita a respetiva referência como erro.

- Só o ficheiro principal deve estar aberto no IBM SPSS Statistics e todos os outros fechados e arquivados na pasta indicada.
- As variáveis chave têm que existir em todos os ficheiros de dados a utilizar no procedimento. Caso tal não aconteça, é reportado um erro tal como indicado na figura abaixo.

4 Extração de carateres

Este menu permite muito facilmente extrair informação de variáveis alfanuméricas, o que só é possível de fazer no IBM SPSS Statistics através de programação em sintaxe.
De uma forma fácil extrai informação com um número fixo ou variável de carateres, com uma determinada posição e/ou delimitados por determinados carateres. A execução do procedimento é feita exclusivamente pela utilização de caixas de diálogo.

4.1 Funcionalidades

São várias as opções que o utilizador tem para proceder à extração de carateres, incluindo a possibilidade de, a partir da informação existente numa variável, criar múltiplas variáveis. Todas as opções de seleção serão descritas adiante.
Tudo o que tem a fazer é indicar a variável donde pretende extrair a informação e qual a operação a realizar.
O resultado que se obtém após executar este procedimento é a criação no ficheiro de dados em utilização de novas variáveis com a informação extraída de uma variável existente.

4.2 Utilização

Para exemplificar a utilização deste procedimento vamos utilizar diferentes ficheiros de dados com pouca informação mas ilustrativos de cada uma das operações.

Extração de carateres

Lista de variáveis disponíveis:

Código Postal [cod_post]

Variáveis selecionadas:

Tipo de operação

Extrair N carateres a partir de uma posição

Número de carateres a extrair

Extrair a partir de posição

Extrair N carateres do fim

Extrair carateres do início até encontrar o carater

Especifique carater ou padrão

Extrair N carateres a partir do carater

Número de carateres a extrair

Especifique carater ou padrão

Extrair tudo entre dois carateres

Especifique primeiro e segundo carater

Extrair blocos separados pelo carater

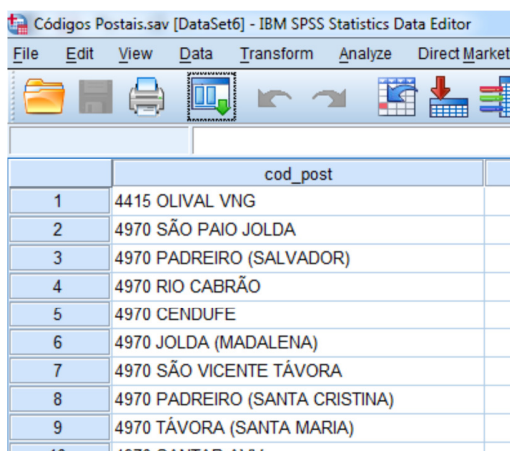
Especifique carater que separa os blocos

Descrição de cada uma das operações possíveis :
- Extrair N carateres a partir de uma posição: Indique o número de carateres a extrair e a posição a partir da qual quer iniciar aquela contagem.
- Extrair N carateres do fim: Indique o número de carateres a extrair sendo que a contagem inicia-se no último carater do registo.
- Extrair carateres do início até encontrar o carater: Indique qual o carater que, ao ser encontrado, termina o processo de extração. Serão extraídos os carateres iniciais do registo até encontrar o carater especificado.
- Extrair N carateres a partir do carater: Indique o número de carateres a extrair e o carater que, encontrado, inicia essa contagem.
- Extrair tudo entre dois carateres: Indique qual o carater, que encontrado, inicia a extração e o carater, que encontrado, termina a extração. Os carateres delimitados por esses carateres serão extraídos.
- Extrair blocos separados pelo carater: Indique o carater que, ao ser encontrado, termina e reinicia o processo de extração. Esta opção permite criar até um máximo de 10 novas variáveis a partir da informação contida numa única variável.

Copyright

Aplicação desenvolvida pela PSE - Produtos e Serviços de Estatística, Lda (www.pse.pt) e licenciada para utilização na empresa PSE.

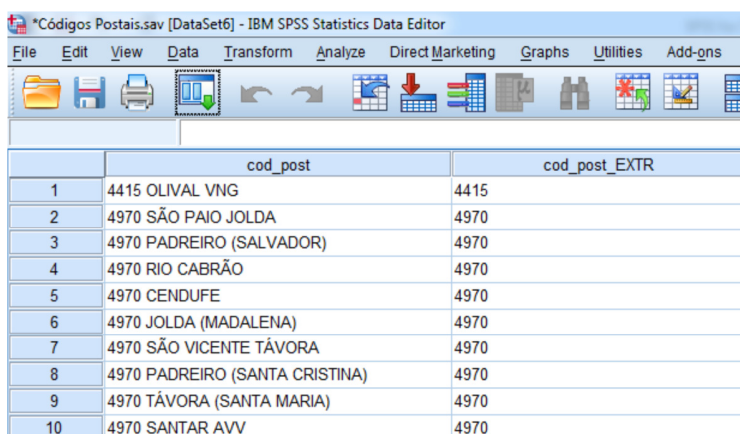
OK | Paste | Reset | Cancel | Help

Em **Lista de variáveis disponíveis** serão identificadas todas as variáveis alfanuméricas disponíveis no seu ficheiro de dados.

Em **Variáveis selecionadas** deverão ser colocadas as variáveis (pode ser mais do que uma) de onde se pretende extrair informação.

Em **Tipo de operação** o utilizador pode optar pelas seguintes opções de extração de carateres:

- Extrair N carateres a partir de uma posição – Nesta operação o utilizador apenas necessita de indicar o número de carateres a extrair e a posição (número de carater) a partir da qual se iniciará a extração.
  Nota: A posição indicada contará para a extração.

No exemplo abaixo o objetivo é extrair os **4** carateres numéricos da variável código postal (cod_post). A extração será iniciada na primeira posição.

	cod_post
1	4415 OLIVAL VNG
2	4970 SÃO PAIO JOLDA
3	4970 PADREIRO (SALVADOR)
4	4970 RIO CABRÃO
5	4970 CENDUFE
6	4970 JOLDA (MADALENA)
7	4970 SÃO VICENTE TÁVORA
8	4970 PADREIRO (SANTA CRISTINA)
9	4970 TÁVORA (SANTA MARIA)

O resultado final será a criação de uma nova variável – cod_post(nome da variável original)_EXTR – contendo os carateres de interesse.

	cod_post	cod_post_EXTR
1	4415 OLIVAL VNG	4415
2	4970 SÃO PAIO JOLDA	4970
3	4970 PADREIRO (SALVADOR)	4970
4	4970 RIO CABRÃO	4970
5	4970 CENDUFE	4970
6	4970 JOLDA (MADALENA)	4970
7	4970 SÃO VICENTE TÁVORA	4970
8	4970 PADREIRO (SANTA CRISTINA)	4970
9	4970 TÁVORA (SANTA MARIA)	4970
10	4970 SANTAR AVV	4970

- Extrair N carateres do fim – Nesta operação o utilizador pretende extrair os últimos N carateres das variáveis selecionadas.

No exemplo abaixo o objetivo é extrair os 3 últimos dígitos da variável CÓDIGO_POSTAL.

	DISTRITO	MUNICIPIO	CODINE	FREGUESIA	CODIGO_POSTAL
1	AVEIRO	ÁGUEDA	010101	Agadão	3750-000
2	AVEIRO	ÁGUEDA	010102	Aguada de Baixo	3750-031
3	AVEIRO	ÁGUEDA	010103	Aguada de Cima	3750-041
4	AVEIRO	ÁGUEDA	010104	Águeda	3750-000
5	AVEIRO	ÁGUEDA	010105	Barrô	3750-351
6	AVEIRO	ÁGUEDA	010106	Belazaima do Chão	3750-362
7	AVEIRO	ÁGUEDA	010120	Borralha	3750-000
8	AVEIRO	ÁGUEDA	010107	Castanheira do Vouga	3750-373
9	AVEIRO	ÁGUEDA	010108	Espinhel	3750-403
10	AVEIRO	ÁGUEDA	010109	Fermentelos	3750-000

É criada a variável CODIGO_POSTAL_EXTR com os carateres pretendidos.

Repare que a designação da nova variável tem sempre a designação da variável original acrescida dos carateres _EXTR.

	DISTRITO	MUNICIPIO	CODINE	FREGUESIA	CODIGO_POSTAL	CODIGO_POSTAL_EXTR
1	AVEIRO	ÁGUEDA	010101	Agadão	3750-000	000
2	AVEIRO	ÁGUEDA	010102	Aguada de Baixo	3750-031	031
3	AVEIRO	ÁGUEDA	010103	Aguada de Cima	3750-041	041
4	AVEIRO	ÁGUEDA	010104	Águeda	3750-000	000
5	AVEIRO	ÁGUEDA	010105	Barrô	3750-351	351
6	AVEIRO	ÁGUEDA	010106	Belazaima do Chão	3750-362	362
7	AVEIRO	ÁGUEDA	010120	Borralha	3750-000	000
8	AVEIRO	ÁGUEDA	010107	Castanheira do Vouga	3750-373	373
9	AVEIRO	ÁGUEDA	010108	Espinhel	3750-403	403
10	AVEIRO	ÁGUEDA	010109	Fermentelos	3750-000	000

- <u>Extrair carateres do início até encontrar o carater</u> – Com esta opção serão extraídos todos os carateres desde o início até que seja identificado o delimitador especificado pelo analista.
  O delimitador é parte integrante da nova variável.

No exemplo abaixo queremos extrair todos os carateres desde o início até ser encontrado o carater "-".

É criada a variável VAR_EXTR:

- <u>Extrair N carateres a partir do carater</u> – Com esta seleção todos os carateres contidos após o delimitador especificado, serão extraídos. O delimitador também contará como carater.

No próximo exemplo pretendemos extrair os seis carateres delimitados por “**-**“.

	VAR
1	161f6569-83c9-449c-bb1d-8d479a46aa31
2	161f6569-83c9-449c-bb1d-8d479a46aa31
3	161f6569-83c9-449c-bb1d-8d479a46aa31
4	161f6569-83c9-449c-bb1d-8d479a46aa31
5	497907b3-dd66-40fe-aa29-d7c0d62d4ef0
6	2184b7f6-dece-4e18-8e4e-4627f46e1486
7	2184b7f6-dece-4e18-8e4e-4627f46e1486
8	2184b7f6-dece-4e18-8e4e-4627f46e1486
9	2184b7f6-dece-4e18-8e4e-4627f46e1486
10	2184b7f6-dece-4e18-8e4e-4627f46e1486

É criada a variável VAR_EXTR.

	VAR	VAR_EXTR
1	161f6569-83c9-449c-bb1d-8d479a46aa31	-83c9-
2	161f6569-83c9-449c-bb1d-8d479a46aa31	-83c9-
3	161f6569-83c9-449c-bb1d-8d479a46aa31	-83c9-
4	161f6569-83c9-449c-bb1d-8d479a46aa31	-83c9-
5	497907b3-dd66-40fe-aa29-d7c0d62d4ef0	-dd66-
6	2184b7f6-dece-4e18-8e4e-4627f46e1486	-dece-
7	2184b7f6-dece-4e18-8e4e-4627f46e1486	-dece-
8	2184b7f6-dece-4e18-8e4e-4627f46e1486	-dece-
9	2184b7f6-dece-4e18-8e4e-4627f46e1486	-dece-
10	2184b7f6-dece-4e18-8e4e-4627f46e1486	-dece-

- Extrair tudo entre dois carateres – Opção que permite extrair os carateres contidos entre dois carateres delimitadores.

Num ficheiro que descreve os distritos, os concelhos e as freguesias, existe uma variável que conjuga os códigos de cada uma dessas variáveis identificando cada um com 2 carateres. Cada código está separado elo carater "-". Ou seja na variável CODIGOS as duas primeiras posições são o código do distrito, as duas posições intermédias correspondem ao código dos concelhos e, finalmente, as duas últimas posições correspondem ao código da freguesia.

Pretende-se extrair os códigos referentes aos concelhos (duas posições intermédias).

Para tal, recorre-se à opção de extração de todos os carateres entre dois delimitadores. Neste contexto, o delimitador correspondem ao mesmo **"-"**

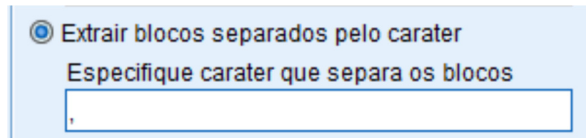
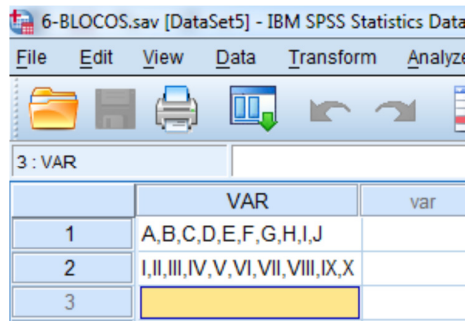
É criada a variável CODIGOS_EXTR.

*5-Tudo entre carateres.sav [DataSet15] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Help Window SPSS for S

1 : CODIGOS_EXTR 01

	DISTRITO	CONCELHO	FREGUESIA	CODIGOS	CODIGOS_EXTR
16	AVEIRO	ÁGUEDA	SEGADÃES	01-01-16	01
17	AVEIRO	ÁGUEDA	TRAVASSÔ	01-01-17	01
18	AVEIRO	ÁGUEDA	TROFA	01-01-18	01
19	AVEIRO	ÁGUEDA	VALONGO DO VOUGA	01-01-19	01
20	AVEIRO	ÁGUEDA	BORRALHA	01-01-20	01
21	AVEIRO	ALBERGARIA-A-VELHA	ALBERGARIA-A-VELHA	01-02-01	02
22	AVEIRO	ALBERGARIA-A-VELHA	ALQUERUBIM	01-02-02	02
23	AVEIRO	ALBERGARIA-A-VELHA	ANGEJA	01-02-03	02
24	AVEIRO	ALBERGARIA-A-VELHA	BRANCA	01-02-04	02
25	AVEIRO	ALBERGARIA-A-VELHA	FROSSOS	01-02-05	02
26	AVEIRO	ALBERGARIA-A-VELHA	RIBEIRA DE FRÁGUAS	01-02-06	02
27	AVEIRO	ALBERGARIA-A-VELHA	SÃO JOÃO DE LOURE	01-02-07	02
28	AVEIRO	ALBERGARIA-A-VELHA	VALMAIOR	01-02-08	02
29	AVEIRO	ANADIA	AMOREIRA DA GÂNDARA	01-03-01	03
30	AVEIRO	ANADIA	ANCAS	01-03-02	03
31	AVEIRO	ANADIA	ARCOS	01-03-03	03
32	AVEIRO	ANADIA	AVELÃS DE CAMINHO	01-03-04	03
33	AVEIRO	ANADIA	AVELÃS DE CIMA	01-03-05	03
34	AVEIRO	ANADIA	MOGOFORES	01-03-06	03
35	AVEIRO	ANADIA	MOITA	01-03-07	03
36	AVEIRO	ANADIA	ÓIS DO BAIRRO	01-03-08	03
37	AVEIRO	ANADIA	SANGALHOS	01-03-09	03
38	AVEIRO	ANADIA	SÃO LOURENÇO DO BAIRRO	01-03-10	03

- Extrair blocos separados pelo carater – Opção muito útil quando se pretende gerar blocos de variáveis. Blocos de variáveis correspondem a novas variáveis (até um máximo de dez) cujo conteúdo corresponderá ao texto especificado na variável de origem e que tem de conter um delimitador. O delimitador fará a separação de informação a ser apresentada nas novas variáveis geradas.

São criadas, neste caso, 10 variáveis com a designação de VAR_EXTR1 ..... VAR_EXTR10.

	VAR	VAR_EXTR1	VAR_EXTR2	VAR_EXTR3	VAR_EXTR4	VAR_EXTR5	VAR_EXTR6	VAR_EXTR7	VAR_EXTR8	VAR_EXTR9	VAR_EXTR10
1	A,B,C,D,E,F,G,H,I,J	A	B	C	D	E	F	G	H	I	J
2	I,II,III,IV,V,VI,VII,VIII,IX,X	I	II	III	IV	V	VI	VII	VIII	IX	X
3											

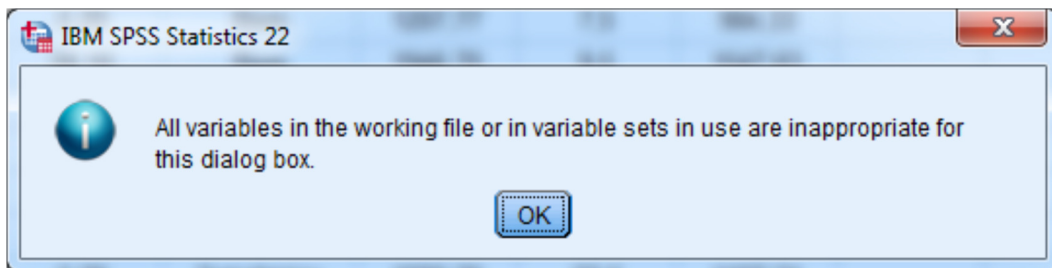
Finalmente, são apresentados cinco botões para efeitos de execução:

- ✓ **OK** – Corresponde à execução do procedimento. Esta opção tem como resultado a criação imediata de novas variáveis no dataset que se encontra ativo.

- ✓ **Paste** – Tal como acontece com os restantes procedimentos de menu do IBM SPSS Statistics, esta opção corresponde à apresentação dos comandos de programação associados ao procedimento de extração de carateres. Esta janela de programação poderá ser guardada e utilizada conjuntamente com outros comandos de programação disponíveis no IBM SPSS Statistics.

- **Reset** – Opção que permite limpar definições tidas no menu.
- **Cancel** – Opção que permite sair do menu.
- **Help** – Opção que permite aceder ao sistema de ajuda associado ao procedimento de extração de carateres.

Considerações

- Apenas são consideradas variáveis alfanuméricas (variáveis em que o campo Type está selecionado como String) para execução de qualquer opção disponível no procedimento de extração de carateres. Se o seu ficheiro de dados não possuir variáveis alfanuméricas aparecer-lhe-á a seguinte mensagem:

- Na opção de “Extrair blocos separados por carater” o procedimento só gerará 10 (dez) novas variáveis. Só pode ser definido um delimitador.
- O carater “&” não se encontra suportado na opção “Extrair blocos separados por carater”.
- O número máximo de carateres apresentando as novas variáveis está limitado a 1000.

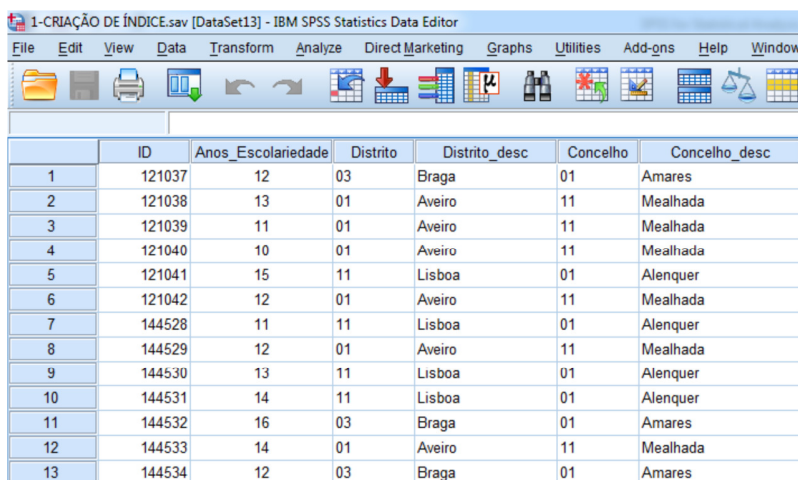
5 Criação de índice

A utilização deste procedimento tem as seguintes vantagens face àqueles que estão disponíveis nativamente com o IBM SPSS Statistics:

- Criação de um índice para numerar sequencialmente todos os casos num ficheiro de dados SPSS.
- Criação de um índice para numerar os casos de acordo com os valores assumidos por uma variável ou um grupo de variáveis. Os casos em que a(s) variável(eis) especificadas têm o mesmo valor, assumem o mesmo número de índice.
- Ordenação dos casos com o mesmo número de índice de acordo com os valores assumidos por uma ou mais variáveis selecionadas.
- Numeração sequencial dos casos que tenham o mesmo número de índice.
- Criação de uma variável que conta o número de casos dentro do mesmo índice

Estas operações são possíveis no IBM SPSS Statistics através de comandos de sintaxe, enquanto que no 'PSE applets for SPSS' obtêm-se através de caixas de diálogo.

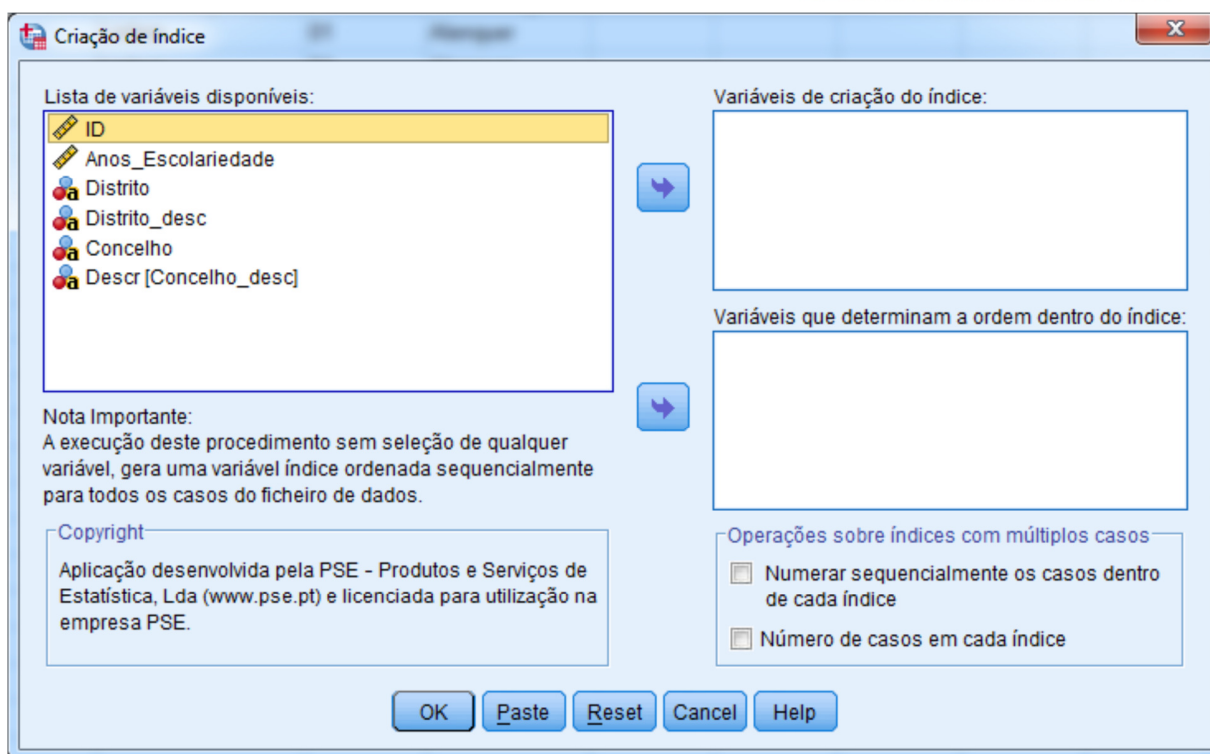
5.1 Funcionalidades

Ao abrir o procedimento, e sem fazer qualquer seleção, se escolher o botão de opção OK é automaticamente criada uma variável designada INDICE que contém a numeração sequencial de todos os casos do ficheiro de dados na ordem em que eles se encontram.
Se colocar uma variável na caixa 'Variáveis de criação do índice' será criado um índice para cada valor dessa variável e cada caso será identificado pelo respetivo índice.
Caso pretenda ordenar os casos com o mesmo índice de acordo com os valores assumidos por uma ou mais variáveis, deve colocar estas na caixa 'Variáveis que determinam a ordem dentro do índice'.
Caso pretenda numerar sequencialmente os casos que tenham o mesmo número de índice basta selecionar a opção 'Numerar sequencialmente os casos dentro de cada índice'.
Caso pretenda obter a informação sobre o número de casos incluídos em cada número de índice deve selecionar a opção 'Número de casos em cada índice'.
A execução deste procedimento tem como resultado a criação de novas variáveis no final do ficheiro de dados conforme as escolhas efetuadas.

5.2 Utilização

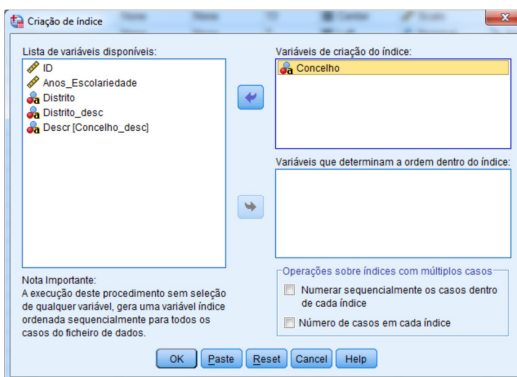
O exemplo, que vamos utilizar para ilustrar este procedimento, apresenta para um conjunto de indivíduos (identificados com a variável ID) os respetivos anos de escolaridade. Para cada caso é também indicado o concelho e o distrito (e respetivos códigos) a que pertence. Com base nesta última informação serão criados as nossas variáveis de índice.

	ID	Anos_Escolariedade	Distrito	Distrito_desc	Concelho	Concelho_desc
1	121037	12	03	Braga	01	Amares
2	121038	13	01	Aveiro	11	Mealhada
3	121039	11	01	Aveiro	11	Mealhada
4	121040	10	01	Aveiro	11	Mealhada
5	121041	15	11	Lisboa	01	Alenquer
6	121042	12	01	Aveiro	11	Mealhada
7	144528	11	11	Lisboa	01	Alenquer
8	144529	12	01	Aveiro	11	Mealhada
9	144530	13	11	Lisboa	01	Alenquer
10	144531	14	11	Lisboa	01	Alenquer
11	144532	16	03	Braga	01	Amares
12	144533	14	01	Aveiro	11	Mealhada
13	144534	12	03	Braga	01	Amares

O procedimento de criação de índice permite a seleção de todo o tipo de variáveis (numéricas, alfanuméricas) com exceção das variáveis de resposta múltipla.

As variáveis que são criadas pelo procedimento são sempre do tipo numérico.

Assim que acede ao menu de criação de índice, pode, sem especificar uma variável obter um índice sequencial de casos (conforme indicação no canto inferior esquerdo).

No ficheiro de dados surgirá uma nova variável – INDICE – que lhe permite identificar, exclusivamente, os casos presentes.

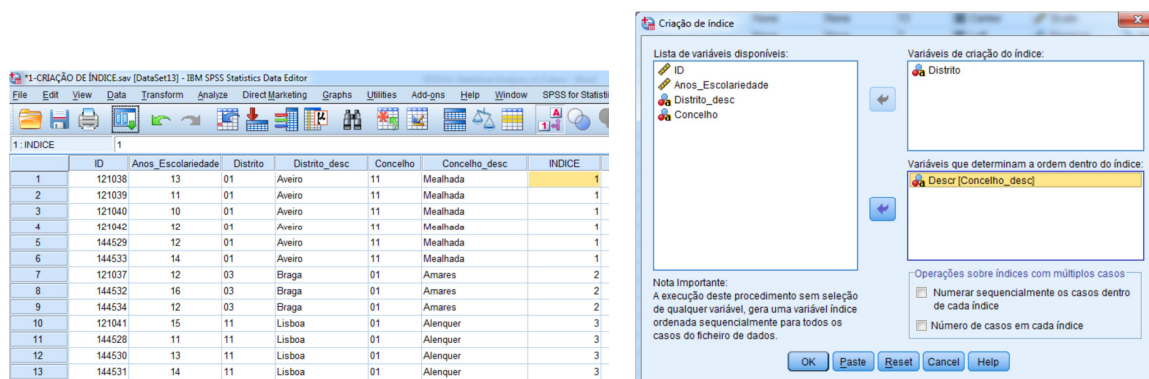
No nosso exemplo o índice assume os valores de 1 a 13 que correspondem aos 13 casos do ficheiro de dados, conforme próxima imagem.

	ID	Anos_Escolariedade	Distrito	Distrito_desc	Concelho	Concelho_desc	INDICE	var
1	121037	12	03	Braga	01	Amares	1	
2	121038	13	01	Aveiro	11	Mealhada	2	
3	121039	11	01	Aveiro	11	Mealhada	3	
4	121040	10	01	Aveiro	11	Mealhada	4	
5	121041	15	11	Lisboa	01	Alenquer	5	
6	121042	12	01	Aveiro	11	Mealhada	6	
7	144528	11	11	Lisboa	01	Alenquer	7	
8	144529	12	01	Aveiro	11	Mealhada	8	
9	144530	13	11	Lisboa	01	Alenquer	9	
10	144531	14	11	Lisboa	01	Alenquer	10	
11	144532	16	03	Braga	01	Amares	11	
12	144533	14	01	Aveiro	11	Mealhada	12	
13	144534	12	03	Braga	01	Amares	13	

Em **Lista de variáveis disponíveis** serão identificadas todas as variáveis disponíveis no seu ficheiro de dados.

Em **Variáveis de criação do índice** deverão ser colocadas todas as variáveis para efeitos de criação de um índice para cada valor que assumam (se selecionar mais que uma variável, cada índice corresponderá a uma combinação dos respetivos valores). Para o caso de incluir duas ou mais variáveis, a criação do índice respeitará a ordem de variáveis colocadas.

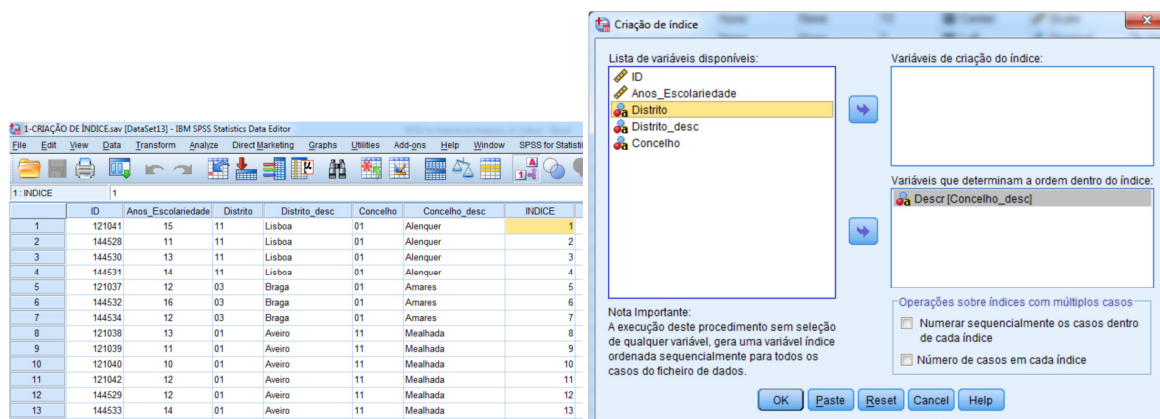
Vejamos os próximos exemplos:

- Especificação de uma variável – Concelho

	ID	Anos_Escolariedade	Distrito	Distrito_desc	Concelho	Concelho_desc	INDICE
1	121041	15	11	Lisboa	01	Alenquer	1
2	144528	11	11	Lisboa	01	Alenquer	1
3	144530	13	11	Lisboa	01	Alenquer	1
4	144531	14	11	Lisboa	01	Alenquer	1
5	121037	12	03	Braga	01	Amares	1
6	144532	16	03	Braga	01	Amares	1
7	144534	12	03	Braga	01	Amares	1
8	121038	13	01	Aveiro	11	Mealhada	2
9	121039	11	01	Aveiro	11	Mealhada	2
10	121040	10	01	Aveiro	11	Mealhada	2
11	121042	12	01	Aveiro	11	Mealhada	2
12	144529	12	01	Aveiro	11	Mealhada	2
13	144533	14	01	Aveiro	11	Mealhada	2

O ficheiro de dados será ordenado pelos códigos dos Concelhos. A variável índice assumirá dois códigos: 1 – respeitante ao concelho (01) - e 2 – respeitante ao concelho (11).

- Especificação de duas variáveis – Concelho e Distrito

O ficheiro de dados será ordenado pelos códigos dos Concelhos (01 e 11) e seguidamente pelos códigos do distrito (01, 03 e 11).

Selecionando uma ou mais variáveis em **Variáveis que determinam a ordem dentro do índice** tem a possibilidade de determinar a ordenação de casos dentro de cada grupo gerado. Para o caso de incluir duas ou mais variáveis, a ordenação do índice respeitará a ordem de variáveis colocadas.

No nosso exemplo, vamos supor que pretendemos que a variável INDICE seja gerada com base nos códigos dos distritos e a sua apresentação seja feita com base numa ordenação por designação dos descritivos.

De referir que caso pretenda especificar apenas informação nesta seção, o procedimento também o permite.

No próximo exemplo, pretendemos a apresentação de resultados com base na descrição dos concelhos.

O índice gerado, corresponderá à ordenação sequencial dos casos (dado que não especificamos variáveis na seção de variáveis a constituírem grupos), mas ordenados de forma alfabética (ascendente) consoante os descritivos assumidos pela variável selecionada (Concelho_desc).

Em **Operações sobre índices com múltiplos casos** o utilizador poderá selecionar uma ou mais das opções disponíveis:

- Numerar sequencialmente os casos dentro de cada índice – Opção que terá como retorno a criação de uma nova variável - *INDICE_SEQUENCIAL* – que listará sequencialmente os casos identificados. A partir do momento que o índice identifica um novo caso, este contador inicia-se.

Como se pode verificar, existem dois códigos de concelho (apresentados nova variável INDICE). A contagem sequencial – variável INDICE_SEQUENCIAL – lista, para o código de concelho "01" os 7 primeiros casos. De seguida, o contador inicia-se (correspondente ao novo código "11" do Concelho, sendo os restantes 6 casos referentes ao segundo concelho identificado.

- Número de casos em cada índice – Para o caso de pretender saber quantos casos pertencem a cada grupo (ou número de índice). Para o efeito, será apresentada a variável - *N_CASOS_GRUPO* – que listará as respetivas ocorrências.

Finalmente, são apresentados cinco botões para efeitos de execução:

- ✓ **OK** – Corresponde à execução do procedimento. Esta opção tem como resultado a criação imediata de novas variáveis no dataset que se encontra ativo.
- ✓ **Paste** – Tal como acontece com os restantes procedimentos de menu do IBM SPSS Statistics, esta opção corresponde à apresentação dos comandos de programação associados ao procedimento de extração de carateres. Esta janela de programação poderá ser guardada e utilizada conjuntamente com outros comandos de programação disponíveis no IBM SPSS Statistics.

- ✓ **Reset** – Opção que permite limpar definições tidas no menu.

- ✓ **Cancel** – Opção que permite sair do menu.
- ✓ **Help** – Opção que permite aceder ao sistema de ajuda associado ao procedimento de criação de índices.

C:\Program Files\IBM\SPSS\Statistics\21\ext\lib\menu4\menu4.htm

Criação de Índice

Terminologia

Casos
Registos em linha de um ficheiro de dados SPSS Statistics. Também conhecidos por observações.
Variáveis
Identificação das colunas na matriz de dados SPSS e que contém informação sobre cada um dos casos.
Dataset
Ficheiro de dados SPSS Statistics.
Lista de variáveis disponíveis
Identificação de variáveis disponíveis no ficheiro de dados e sobre as quais podemos executar o procedimento.
- Neste procedimento não é obrigatório a seleção de qualquer variável.

Variáveis de criação do índice
Identificação das variáveis que foram selecionadas da lista de variáveis disponíveis e cuja combinação de valores determinará a criação do índice:
- Pode selecionar uma ou mais variáveis. A criação do índice respeitará a ordem das variáveis colocados nesta caixa.

Variáveis que determinam a ordem dentro do índice
Identificação das variáveis que vão ordenar de forma ascendente os casos incluídos num mesmo índice
- Pode selecionar uma ou mais variáveis. A ordenação respeitará a ordem das variáveis colocadas nesta caixa.

Operações sobre índices com múltiplos casos
Operações adicionais a realizar com a criação do índice. As opções disponíveis são as seguintes:
- Numerar sequencialmente os casos dentro de cada índice.
Ao selecionar esta opção pretende que os casos que façam parte de um mesmo índice sejam numerados sequencialmente e de forma ascendente. Esta operação cria uma nova variável com a designação de INDICE_SEQUENCIAL.
- Número de casos em cada índice.
Ao selecionar esta opção pretende saber quantos casos pertencem a cada índice. Esta operação cria uma nova variável com a designação N_CASOS_GRUPO.

Utilidade

Este procedimento facilita o processo de criação de índices para todo o ficheiro de dados ou para grupos definidos por variáveis presentes no dataset.
- Quando existe mais do que um caso com o mesmo índice é muito útil a utilização das operações incluidas em 'Operações relativas a índices com múltiplos casos'. Pode numerar sequencialmente cada caso dentro de índice e pode saber quantos casos pertencem a cada índice.
- A obtenção de um índice para todo o ficheiro de dados é feita quando não selecionamos nenhuma variável para a caixa 'Variáveis de criação do índice'.

Funcionamento

Na caixa 'Lista de variáveis disponíveis' podemos selecionar as variáveis cujos valores, simples ou combinados, criarão uma nova variável no ficheiro de dados que se designará INDICE e que, de uma forma sequencial ascendente, numera cada registo. Os registos que pertençam à mesma combinação de valores das variáveis selecionadas terão o mesmo índice. Utilizando o botão com a seta devemos colocá-las na caixa 'Variáveis selecionadas'.

Considerações

- A criação do índice é feita por ordem ascendente dos valores ordenados do ficheiro de dados ou das variáveis selecionadas para definição de grupos.
- Pode gerar, individual ou simultaneamente as duas opções disponíveis em "Operações sobre índices com múltiplos casos", mesmo que nas duas anteriores seções de especificação de variáveis não selecione variáveis. Neste cenário, os casos serão tratados como sendo primários e, consequentemente, o índice gerado será sequencial (identificação exclusiva), ou seja, listará os casos desde o primeiro até ao último.
- Caso no ficheiro de dados já tenham criadas variáveis de índice (através do menu Criação de índice) e pretenda voltar a executar uma outra vez este procedimento (por exemplo, porque tem outro critério definido de criação de índices) alertamos que, por defeito, as variáveis serão recalculadas. Se pretender manter os vários índices criados, sugerimos que atribua novas designações às variáveis criadas pelo procedimento.